



MUSÉUM
NATIONAL D'HISTOIRE NATURELLE



Centre d'Écologie et des
Sciences de la Conservation

VIGIENATURE
Un réseau de citoyens qui fait avancer la science

Ecologie et big data



Biodiversité et crowdsourcing

(monitoring)

(participation)

Romain Julliard

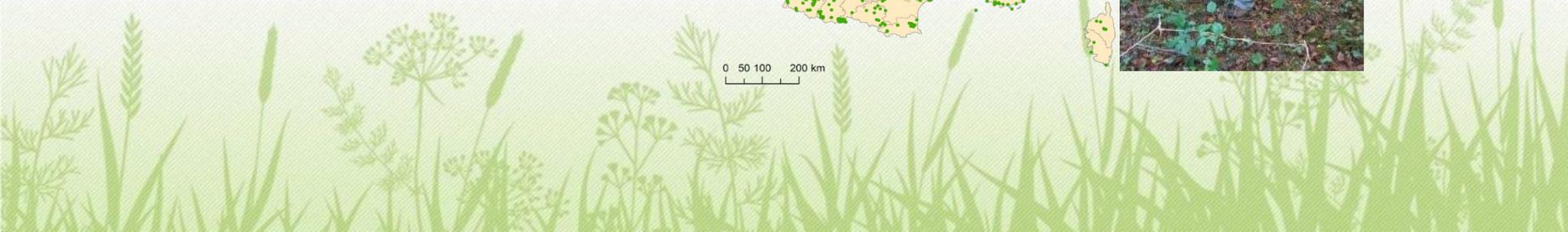
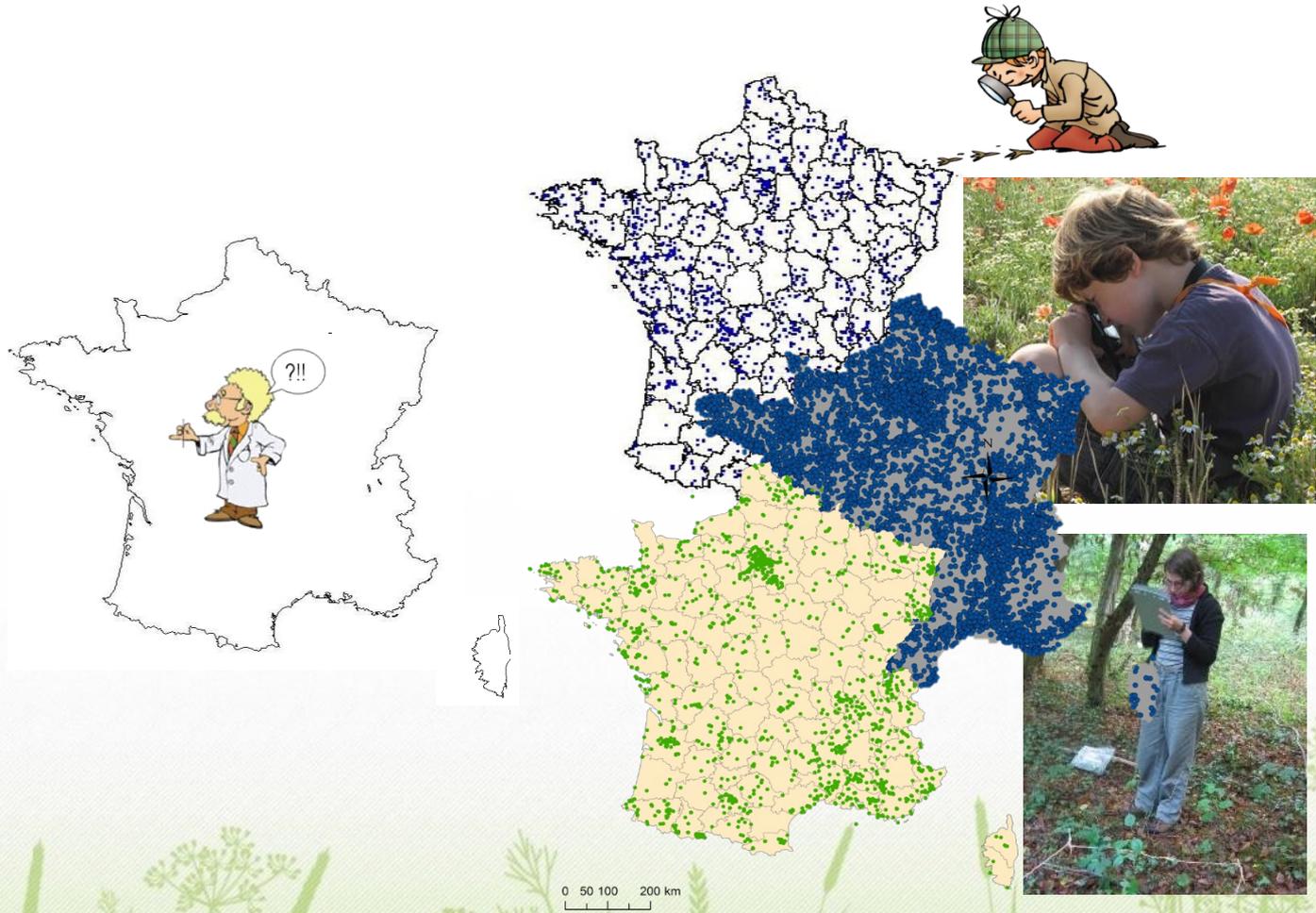
Objectifs de

VIGIENATURE

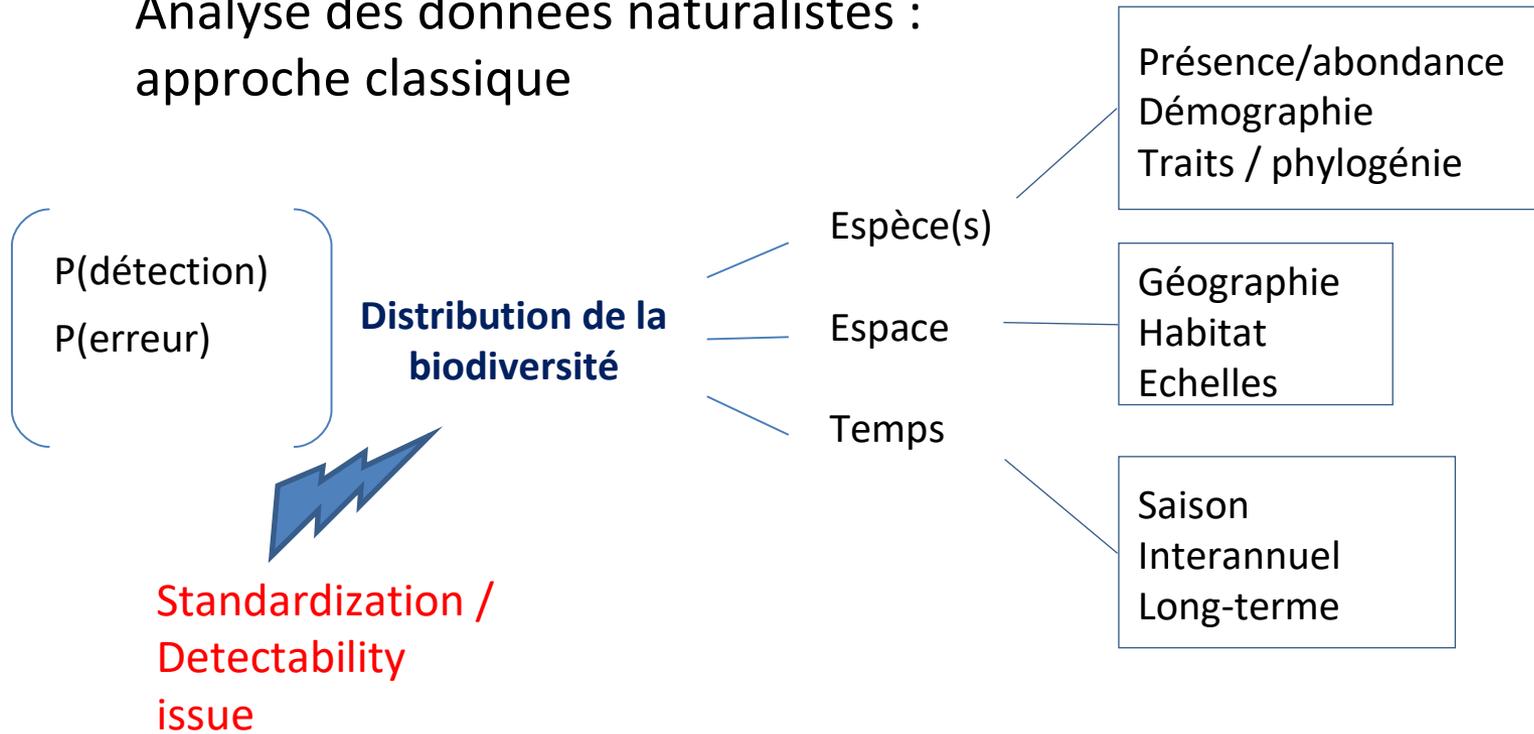
Un réseau de citoyens qui fait avancer la science

Un programme de recherche du Muséum national d'Histoire naturelle

- **surveiller** l'état de la biodiversité,
 - » **Mais les espèces sont très diverses !**
- **diagnostiquer** les causes des variations
 - » **Mais causes multiples !**
- proposer des **scénarios**



Analyse des données naturalistes : approche classique



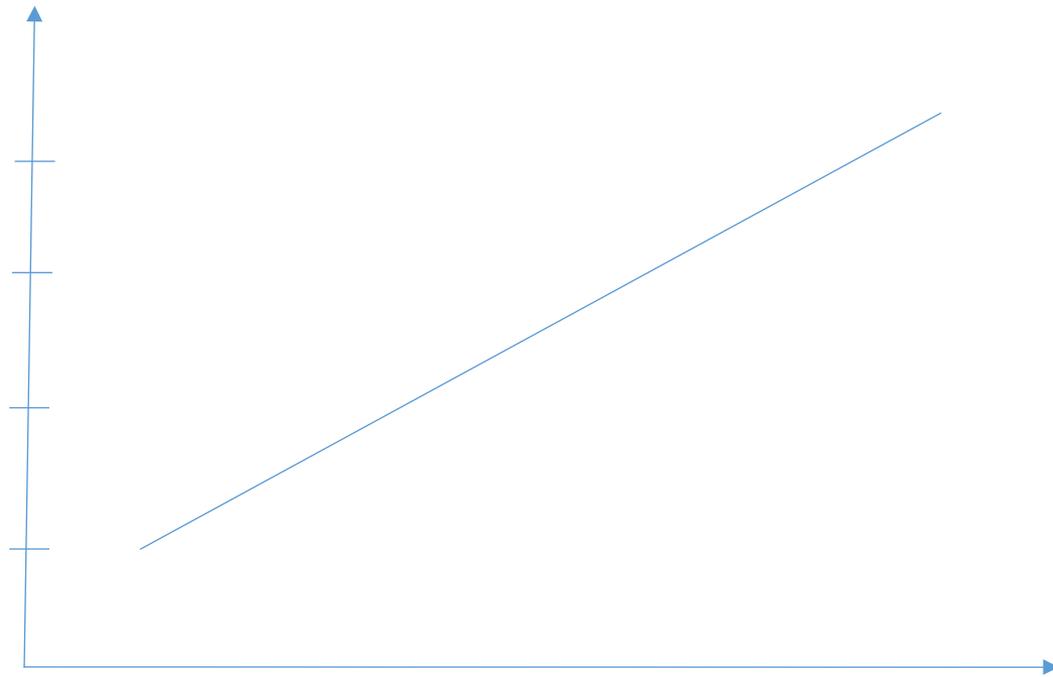
Increasing data
quality (fit for
purpose of
monitoring)

Abundance +
sampling
design

Abundance +
effort

Species check
list

Incidental
data



=> Increasing cost of collecting data

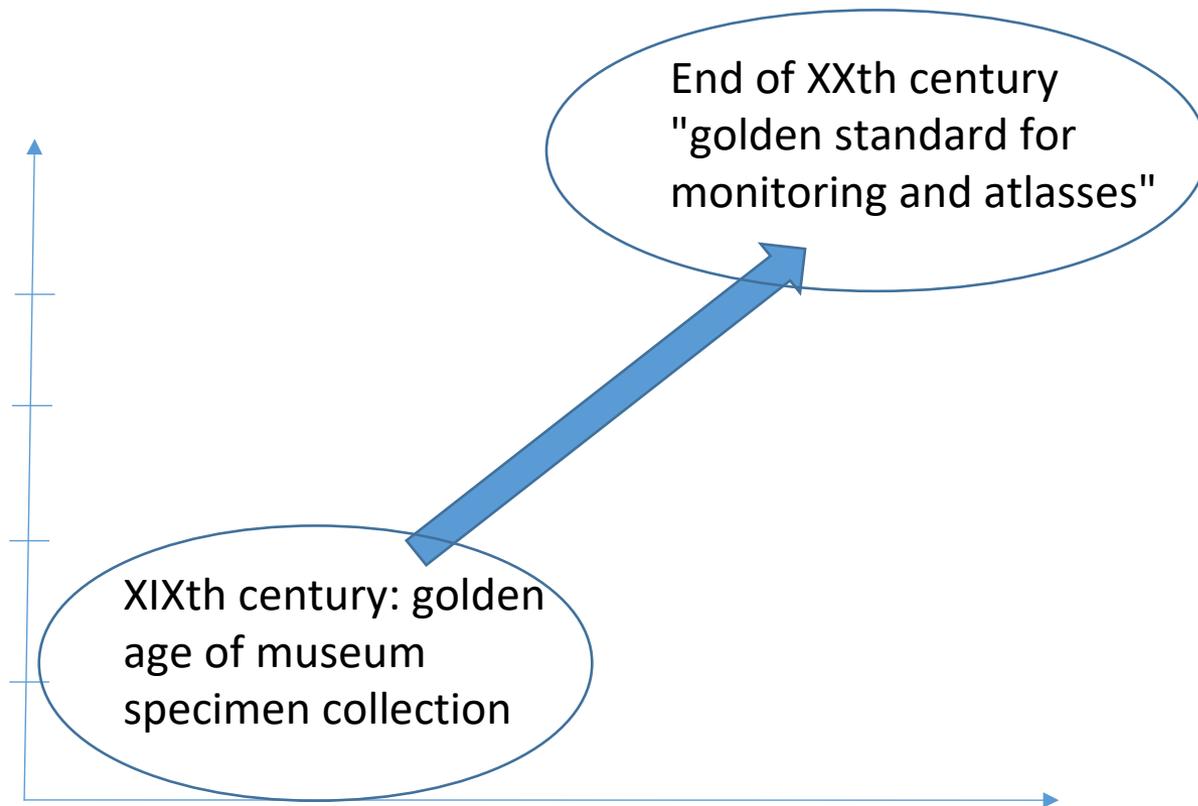
Increasing data quality (fit for purpose of monitoring)

Abundance + sampling design

Abundance + effort

Species check list

Incidental data



End of XXth century
"golden standard for monitoring and atlases"

XIXth century: golden age of museum specimen collection

=> Increasing cost of collecting data

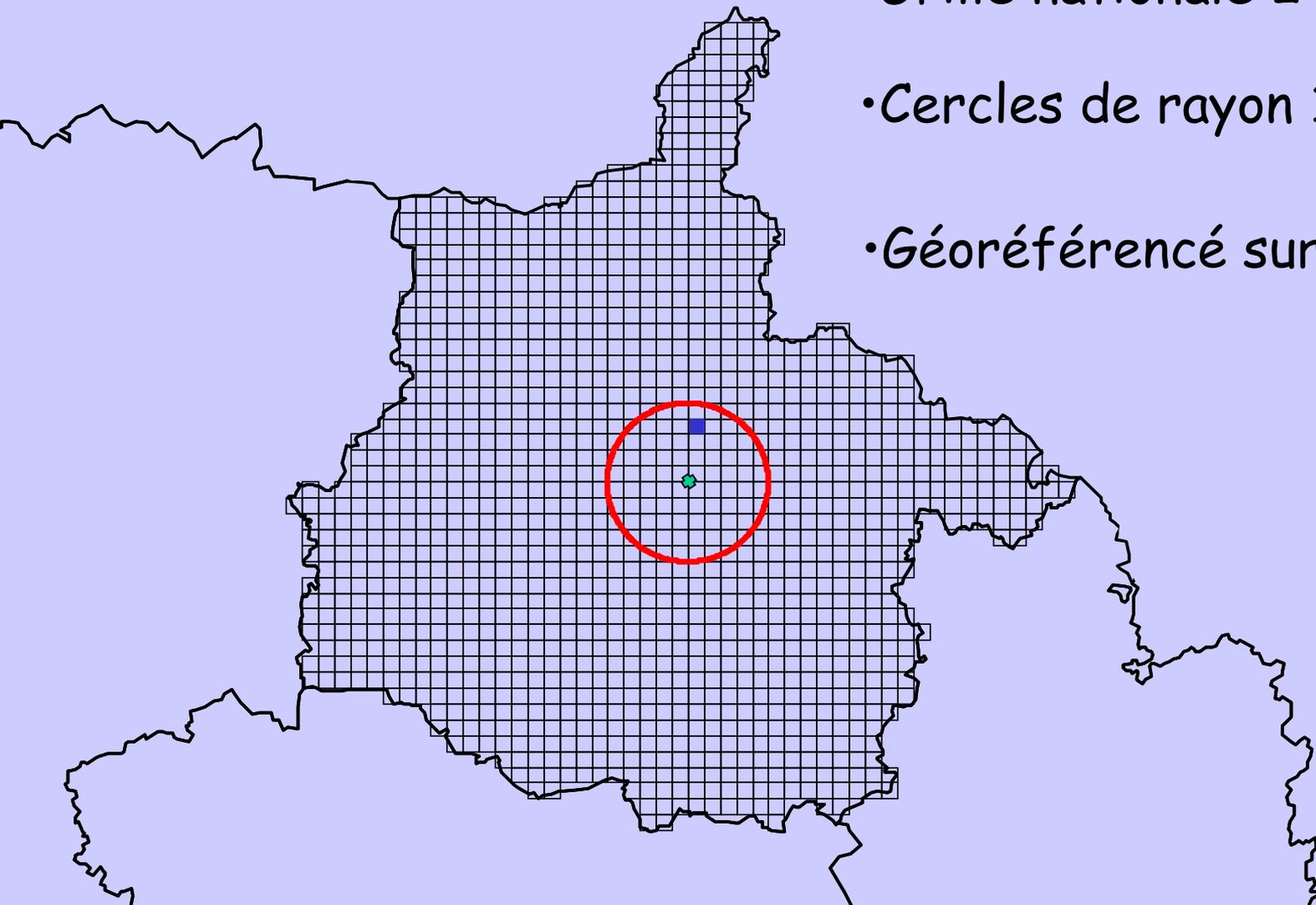
Le programme Suivi Temporel des Oiseaux Communs STOC

Basé sur les ornithologues amateurs
Début : 1989



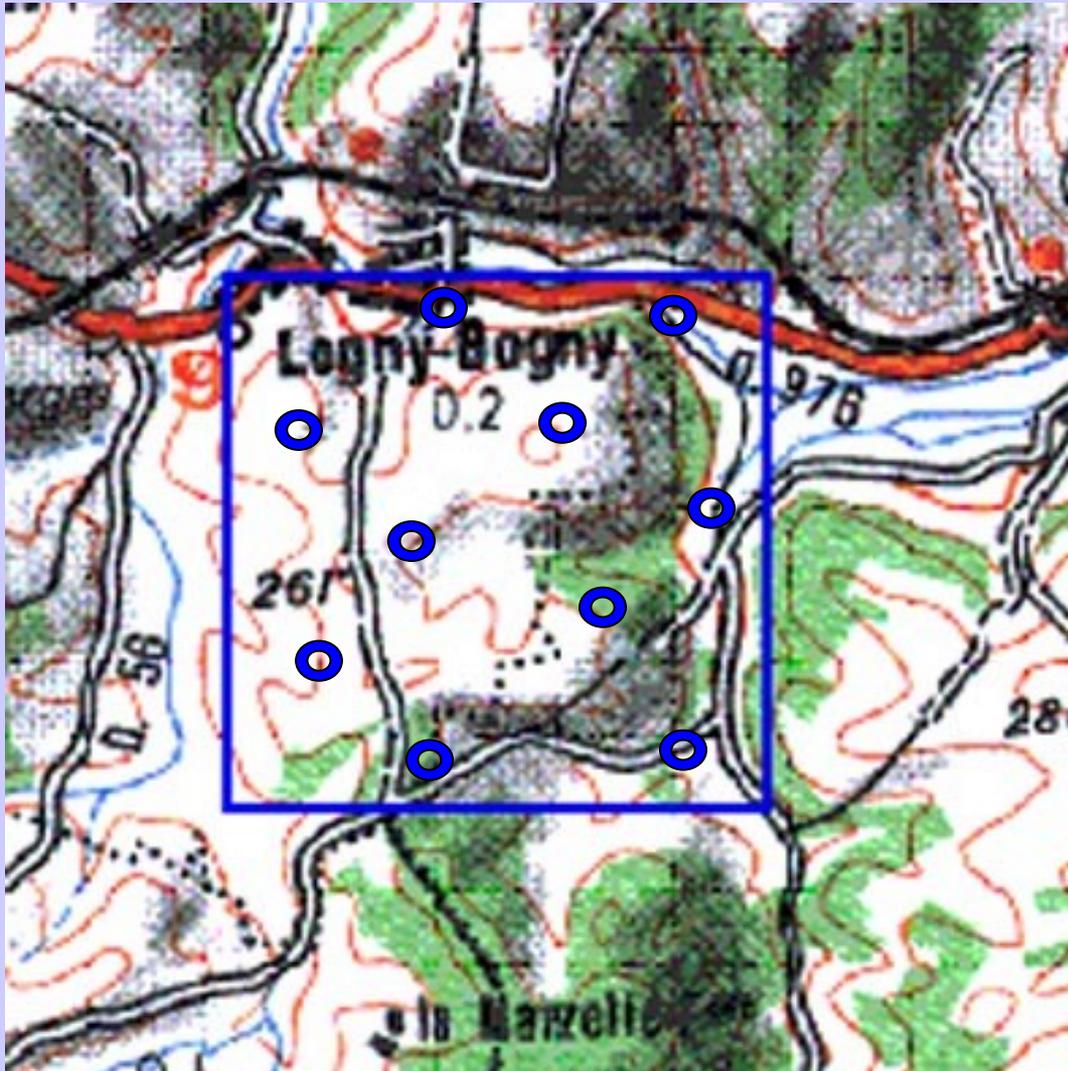
Comment procède-t-on au tirage au sort des sites à prospector ?

- Grille nationale 2 x 2 km
- Cercles de rayon 10 km
- Géoréférencé sur SIG



Distribution homogène des EPS

Proportionnelle aux habitats



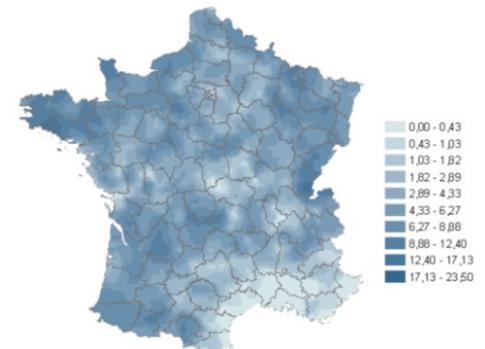
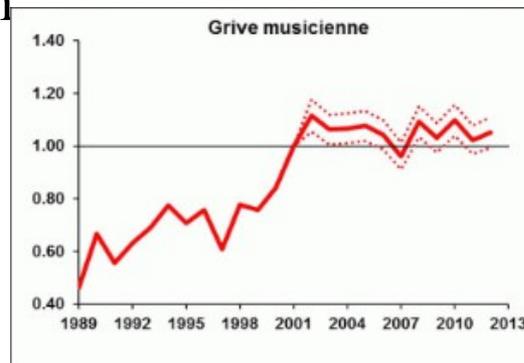
10 EPS dans un carré

Faits deux fois par an
avant et après le 8 mai

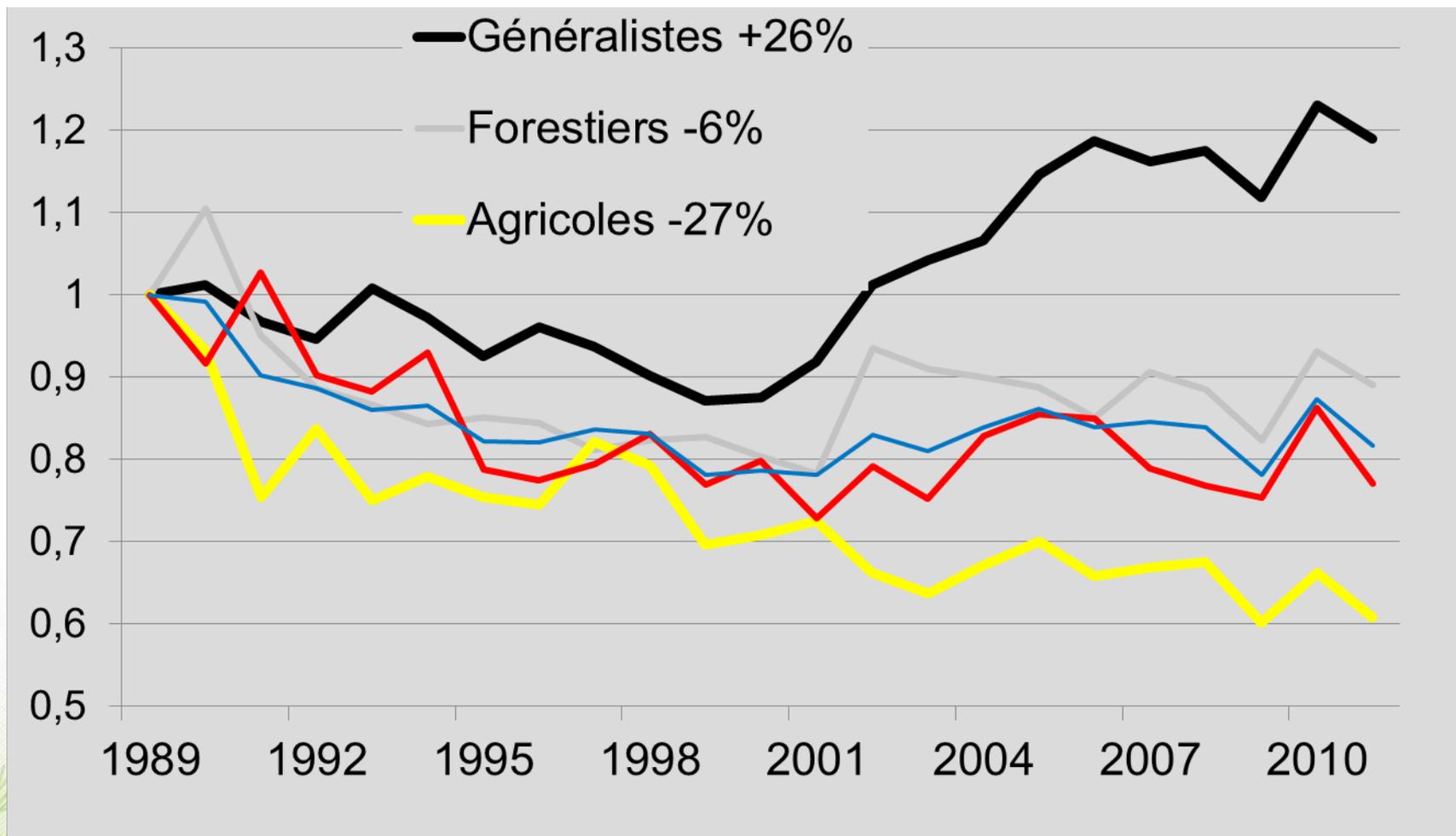
**Description
standardisée de
l'habitat**



- 25 ans
- +/-900 x 10 x 2 relevés/an
- 200k / 250k données/an
- 2,5 M données
- 70 publications



L'indicateur STOC



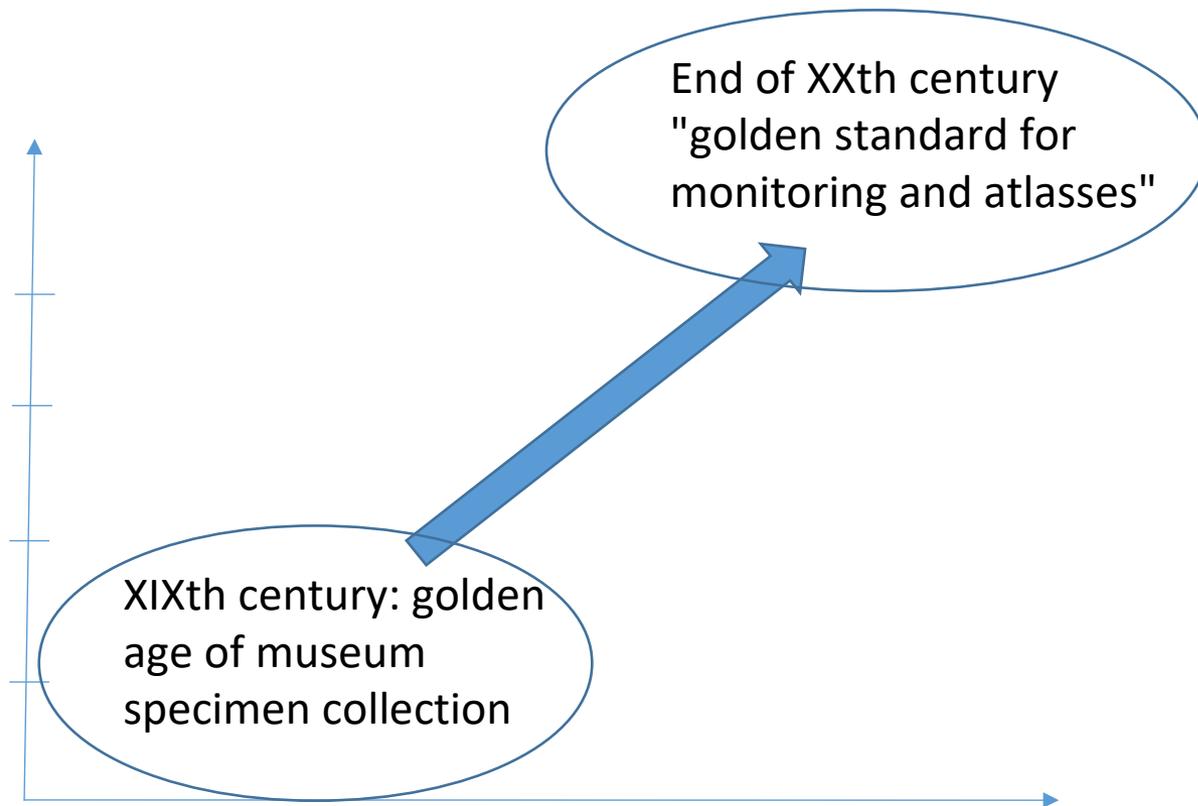
Increasing data quality (fit for purpose of monitoring)

Abundance + sampling design

Abundance + effort

Species check list

Incidental data



End of XXth century
"golden standard for
monitoring and atlases"

XIXth century: golden
age of museum
specimen collection

=> Increasing cost of collecting data

Increasing data quality (fit for purpose of monitoring)

Abundance + sampling design

Abundance + effort

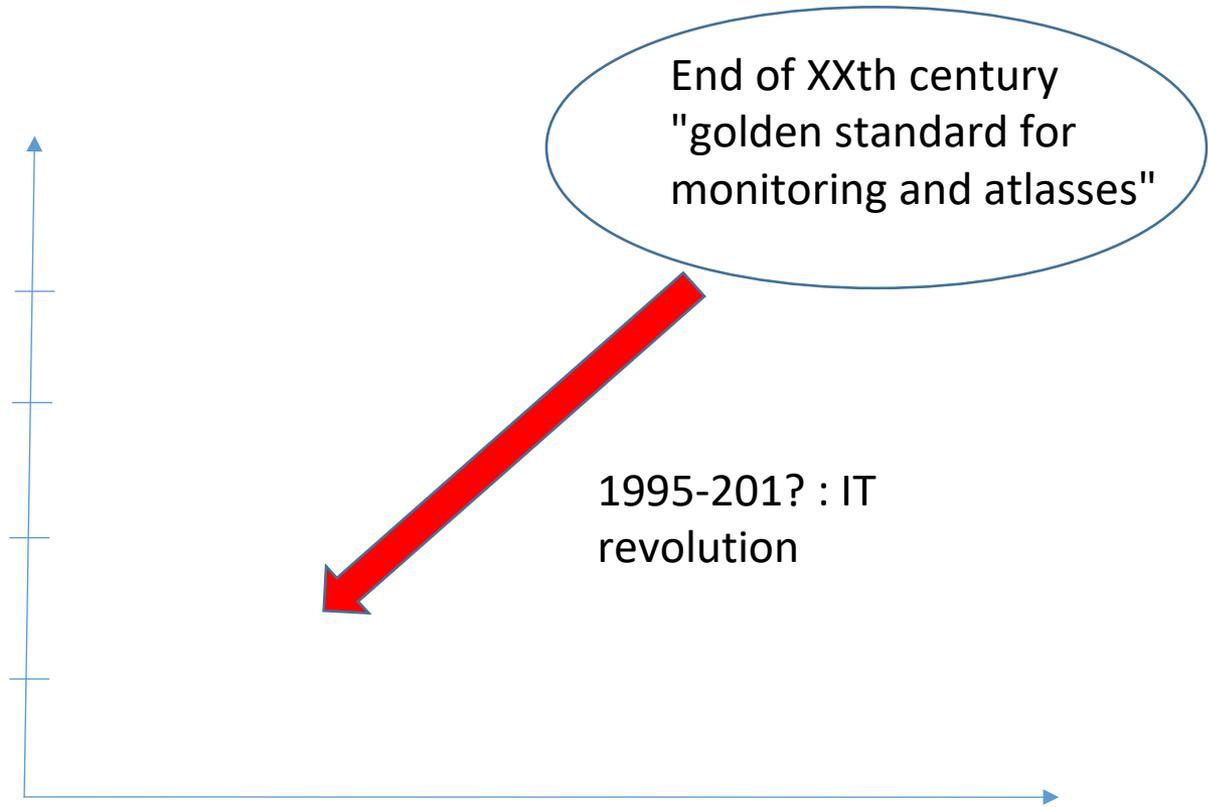
Species check list

Incidental data

End of XXth century
"golden standard for monitoring and atlases"

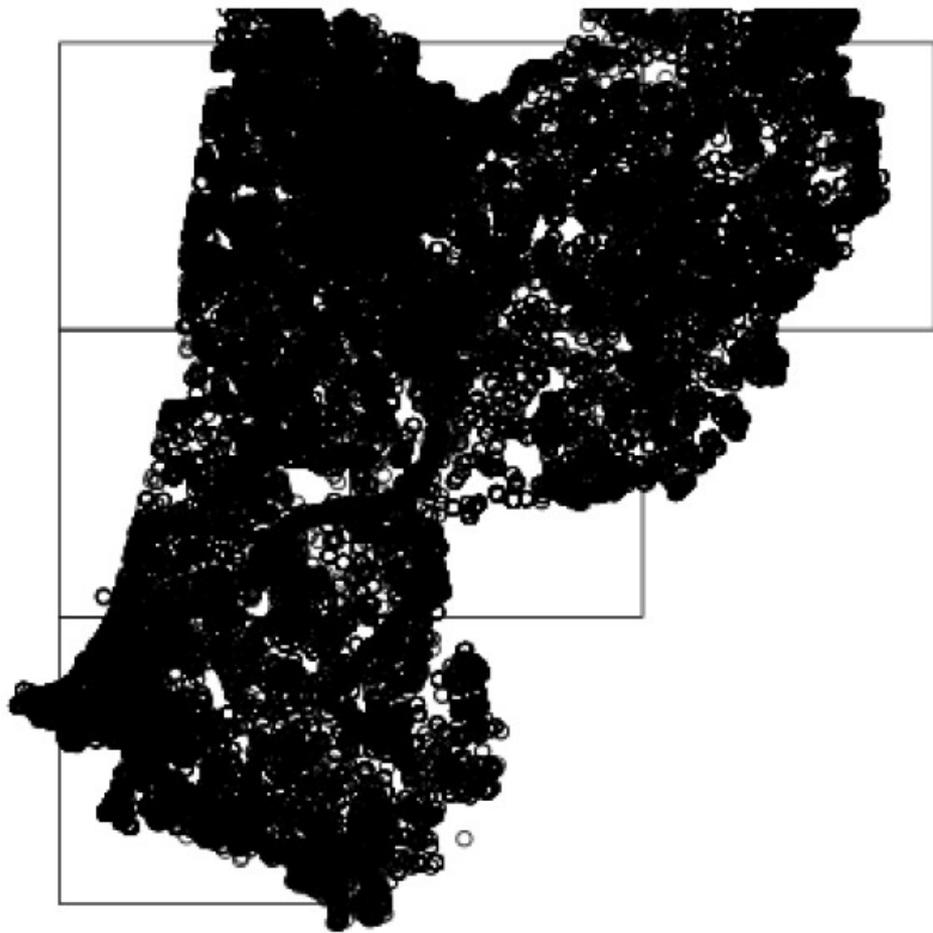
1995-201? : IT revolution

=> Increasing cost of collecting data

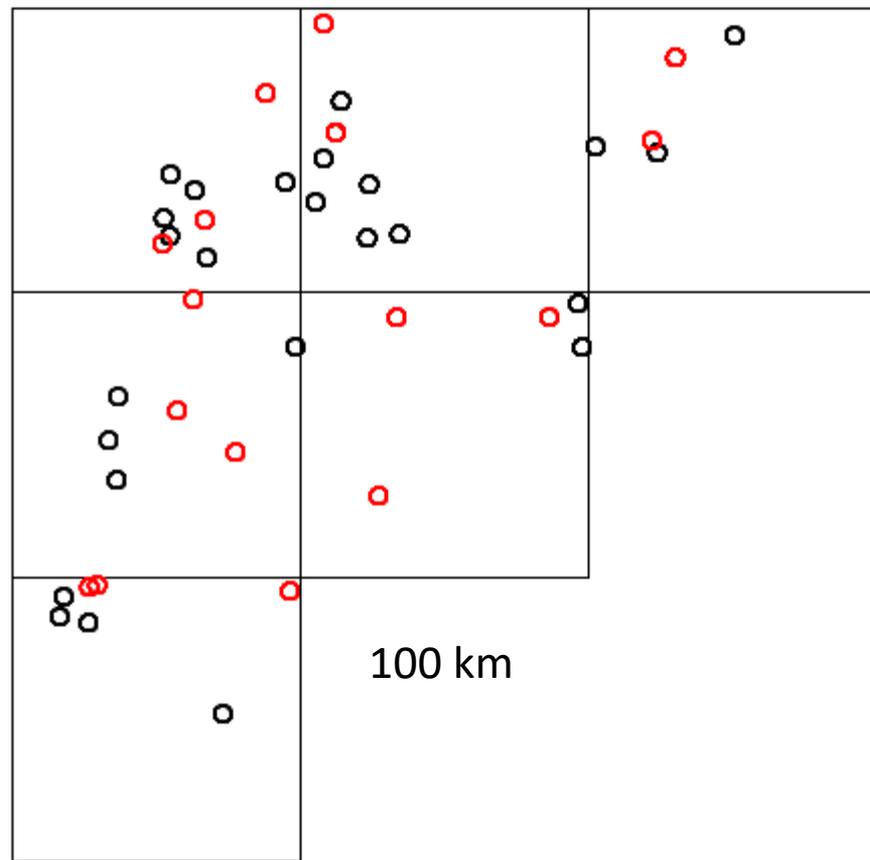


La « data-valanche » touche (enfin) la biodiversité !

- XX^{ème} : pour $n > 30$, loi des grands nombres 😊
- XXI^{ème} : Réseaux d'observateurs + Internet =
bases de données collaborative (des millions de données dites opportunistes ; crowdsourcing)
- (+ virage technologique : la saisie en ligne in situ)

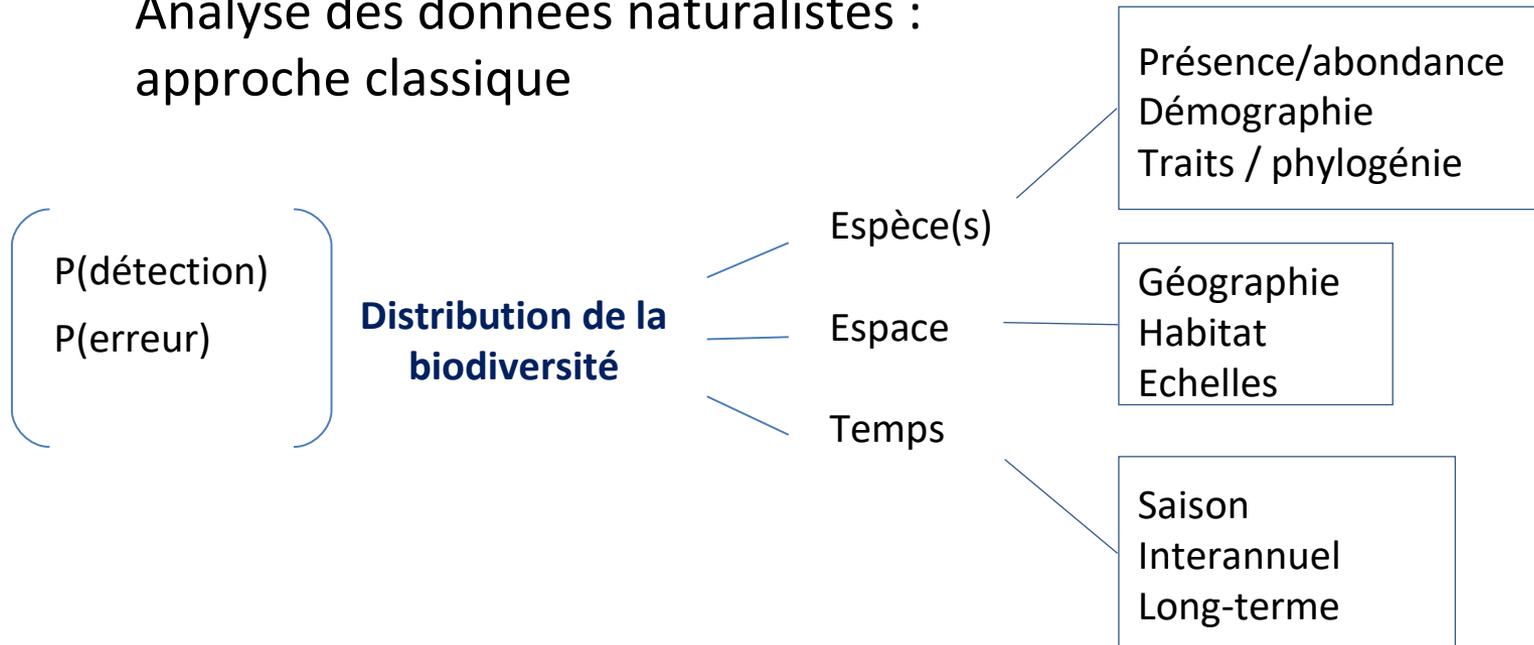


LPO

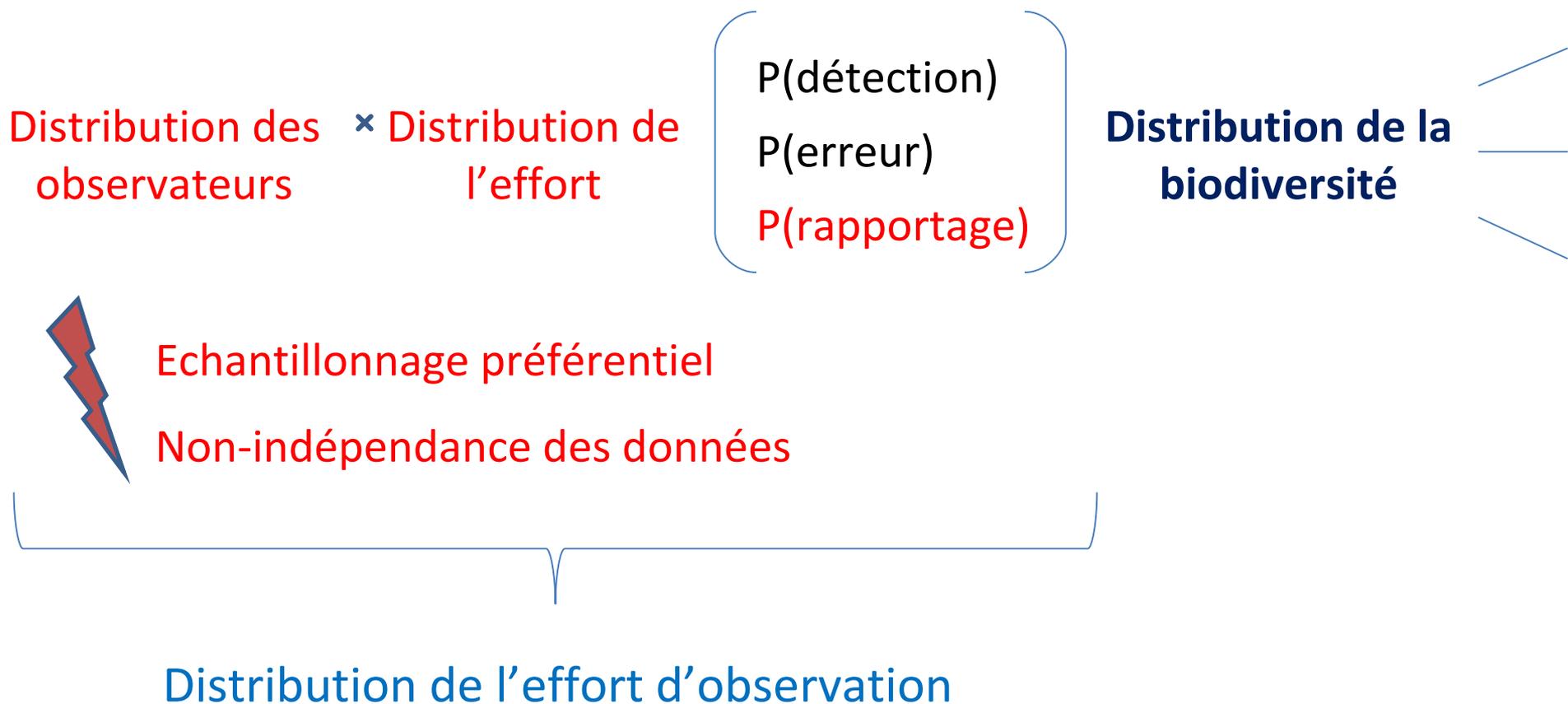


STOC

Analyse des données naturalistes : approche classique



Une observation dans une base de données multi-observateurs c'est le résultat :



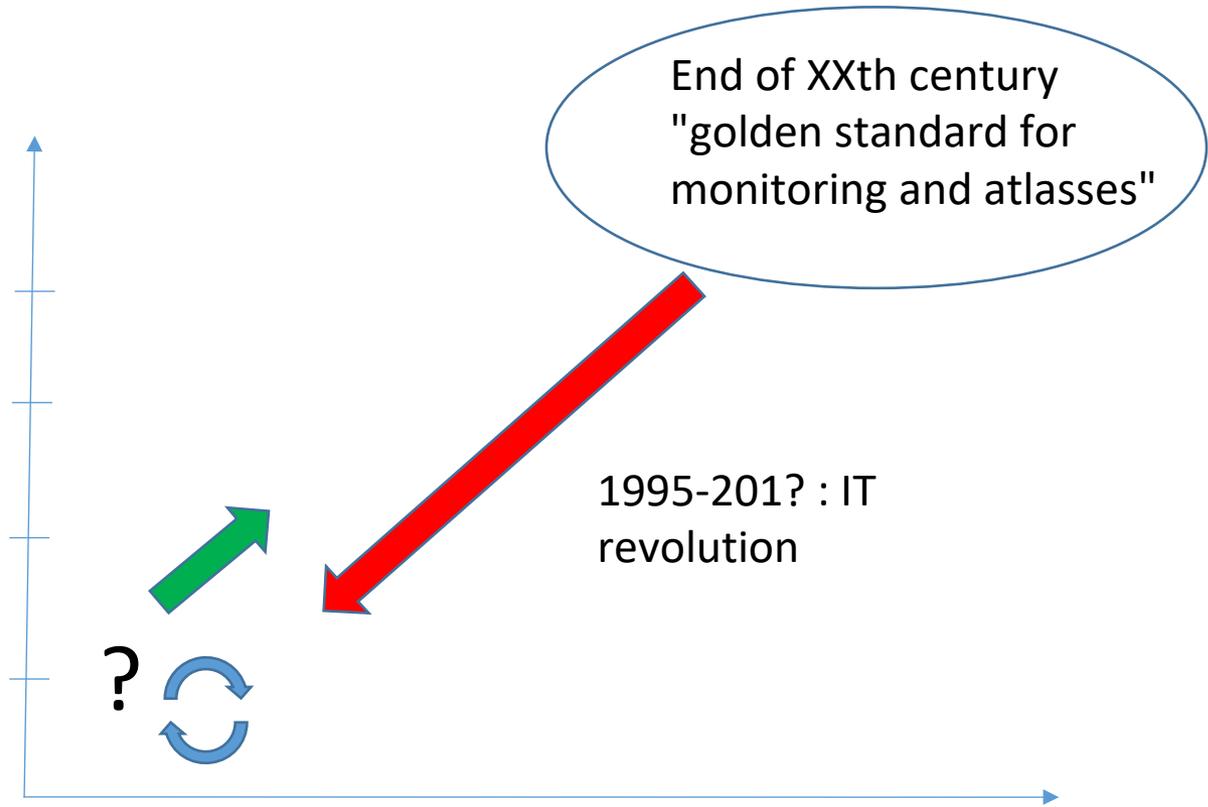
Increasing data quality (fit for purpose of monitoring)

Abundance + sampling design

Abundance + effort

Species check list

Incidental data



End of XXth century
"golden standard for monitoring and atlases"

1995-201? : IT revolution

=> Increasing cost of collecting data

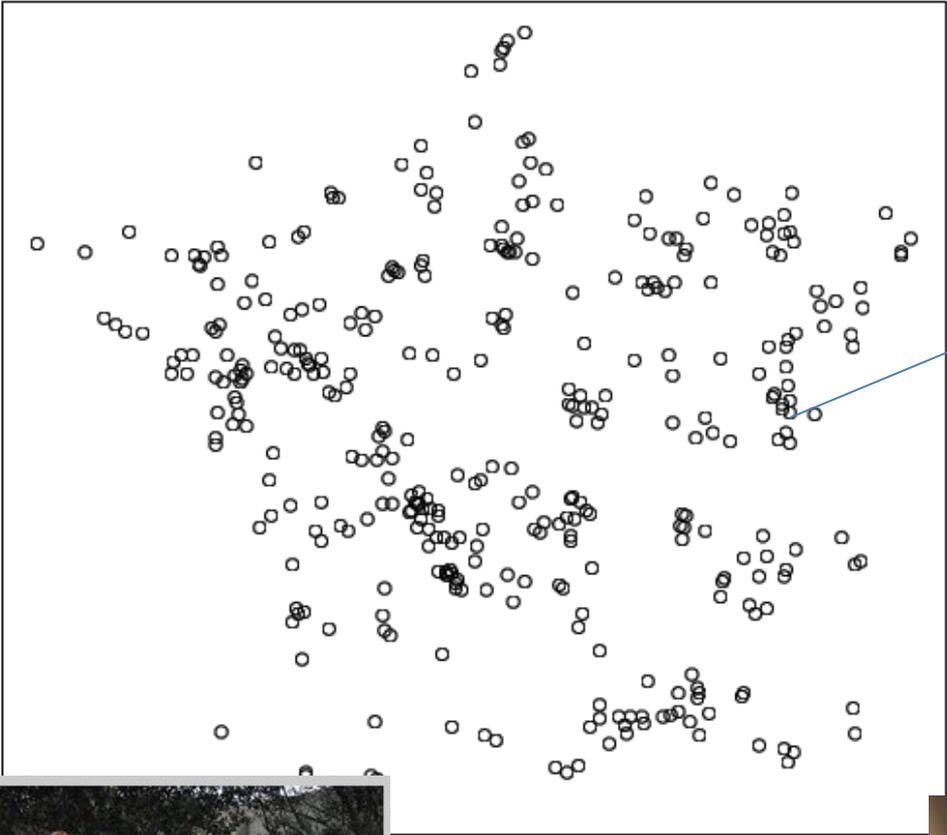
Capitalizing on opportunist data for monitoring biodiversity

- **Christophe Giraud, Camille Coron**, Labo de Math, Orsay
- **Clément Calenge**, ONCFS
- **Romain Julliard**, MNHN-CNRS,

Ecologie

Biométrie

Statistique



$X_{i,j}$ abondance observée de l'espèce i sur le site j



The multi site – multi observer problem :

$$X_{ij} \sim \text{Poisson}(A_{ij} \cdot O_{ij})$$

Observation process

True Abundance

Le problème multi-espèces multi-sites :

$$X_{ij} \sim \text{Poisson}(A_{ij} \cdot E_j \cdot P_i)$$

Probabilité de détection

Effort d'observation

Abondance réelle

=> IJ observations ; $IJ + I + J$ paramètres à estimer 😞

Le problème multi-espèces multi-sites, effort connu :

$$X_{ij} \sim \text{Poisson}(A_{ij} \cdot E_j \cdot P_i)$$

Effort d'observation connu,
à une constante près c

$$X_{ij}/(cE_j) \approx A_{ij}P_i/c \approx \tilde{A}_{ij}$$

=> Species specific relative abundance

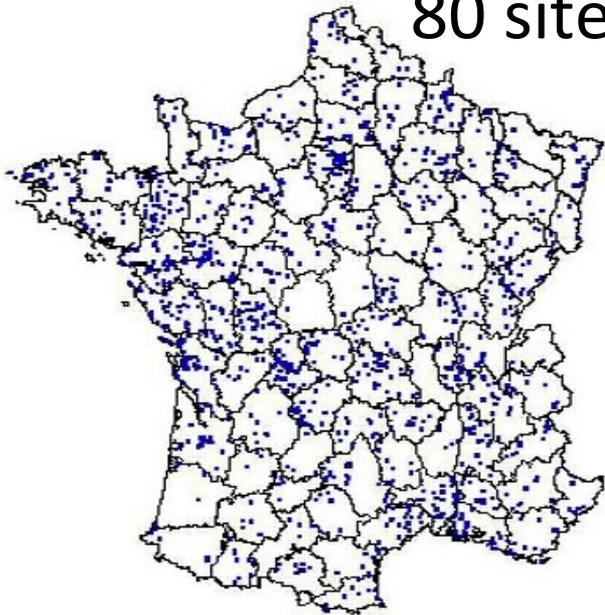
=> IJ observations ; IJ paramètres à estimer

Exemple d'effort connu : suivis standardisés.

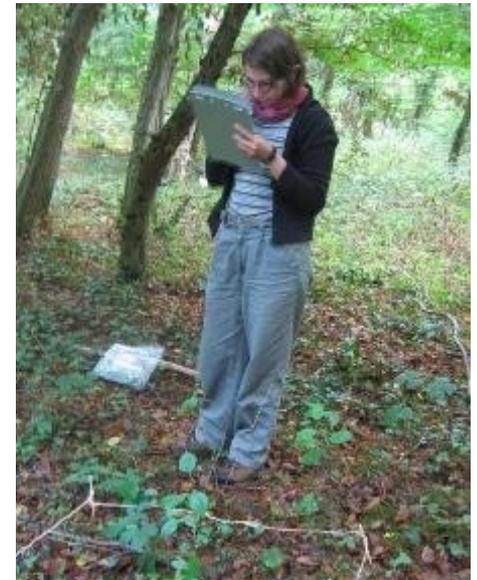


Muséum
national
d'Histoire
naturelle

En France: 1000 sites suivis pour les oiseaux
300 sites suivis pour les chauves-souris
120 sites suivis pour les papillons
80 sites suivis pour les plantes



**Idée : combiner
données standardisées
et données
opportunistes**



Le problème multi-espèces multi-sites multi-protocoles :

$$X_{ijk} \sim \text{Poisson}(A_{ij} \cdot E_{jk} \cdot P_{ik})$$

Probabilité de détection

Effort d'observation

Abondance réelle

Le problème multi-espèces multi-sites multi-protocoles :

$$X_{ijk} \sim \text{Poisson}(A_{ij} \cdot E_{jk} \cdot P_{ik})$$

Probabilité de détection

Effort d'observation

Abondance réelle

=> IJK observations ; IJ + JK + IK paramètres à estimer



Mais ! Pas identifiable...

The multi site – multi observer – multi scheme problem :

$$X_{ijk} \sim \text{Poisson}(A_{ij} \cdot E_{jk} \cdot P_{ik})$$

Species specific detectability
Site specific Observation Effort
True Abundance

=> IJK observations ; IJ + JK + IK parameters 😊
But ! Not all identifiable...

The math tells us:

We need to add $I + J + K - 1$ constraints

We show that if effort is known for scheme $k=0$ (standardized data set), then then E_{jk} pour $k \neq 0$ can be estimated, as well as P_{ik} / P_{i0}

So that the relative abundance \tilde{A}_{ij} can be estimated

But \tilde{A}_{ij} can be estimated with the standardized data set alone! What is the gain?

$$X_{ij} \sim \text{Poisson}(A_{ij} \cdot E_j \cdot P_i)$$

Jeux de données effort connu :
IJ observations ; IJ paramètres à estimer

$$X_{ijk} \sim \text{Poisson}(A_{ij} \cdot E_{jk} \cdot P_{ik})$$

+ 1 jeux de données effort inconnu :
+ IJ observations ; + (I+J-1) paramètres à estimer

- Indeed we show that the precision of A_{ij} estimates is always improved compared to the use of the standardized data set alone
- The gain in precision may be substantial if opportunistic data are abundant

$$\text{var}(\hat{A}_{ij}) = \text{var}(\hat{A}_{ij}^0) \times \frac{P_{i0}A_{ij}}{\sum_l P_{l0}A_{lj}} \times \left(1 + \frac{\gamma_{ij}}{E_{j1}}\right) + O\left(E_{j1}^{-3/2}\right)$$

with

$$\gamma_{ij} = \frac{\sum_{l:l \neq i} P_{l0}^2 A_{lj} + \left(\sum_{l:l \neq i} P_{l0} A_{lj}\right)^2 / A_{ij}}{\sum_l P_{l0} A_{lj}} \times \frac{1 + E_{j0} \sum_l P_{l0} A_{lj}}{\sum_l P_{l0} A_{lj}}.$$

In particular, we have the asymptotic variance reduction factor

$$\frac{\text{var}(\hat{A}_{ij})}{\text{var}(\hat{A}_{ij}^0)} \xrightarrow{E_{j1} \gg 1} \frac{P_{i0}A_{ij}}{\sum_l P_{l0}A_{lj}}.$$

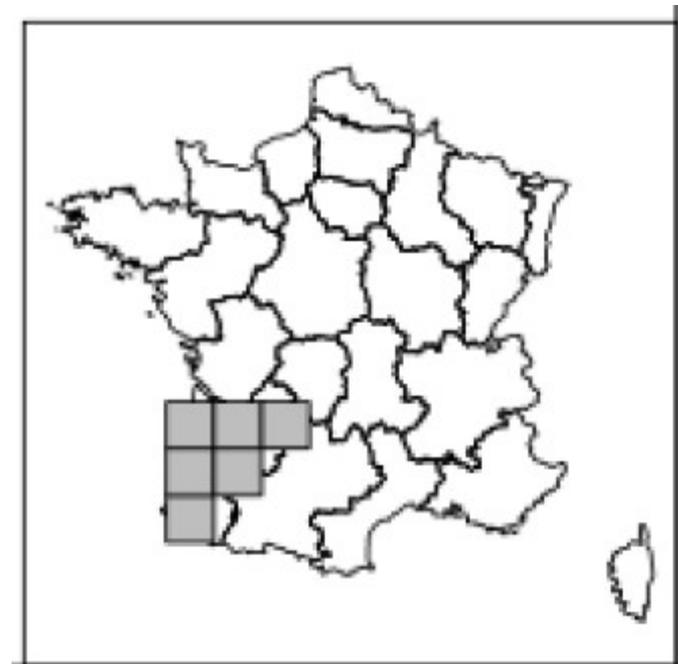
- In addition, we show that, provided that there are some species in common to the standardized and the opportunistic dataset,

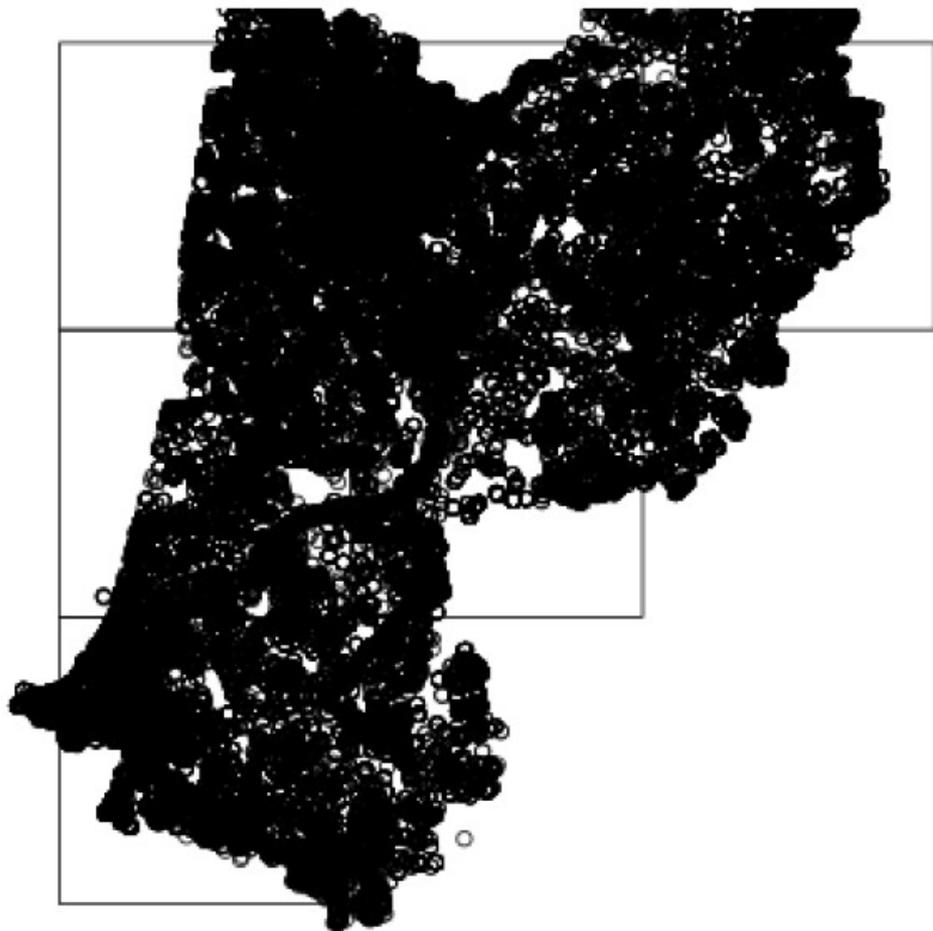
we may get \tilde{A}_{ij} for species only detected in the opportunistic dataset !!

(E_{j1} is estimated from the common set of species; no discontinuity in variance estimation when $X_{i0} = 0$)

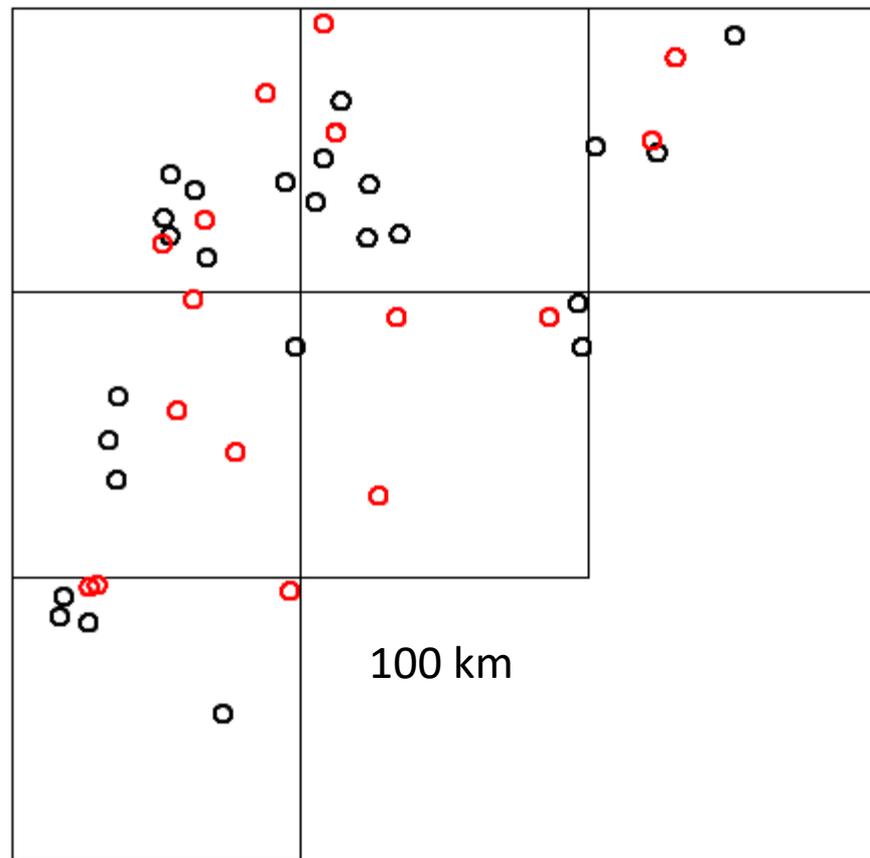
Illustration based on a bird dataset from Aquitaine, south-western France

- Site j = 100x100 km square
- STOC = standardized,
sampling design
- LPO= opportunistic dataset





LPO



STOC1 + STOC2

Pour les 22 espèces représentées dans tous les lots

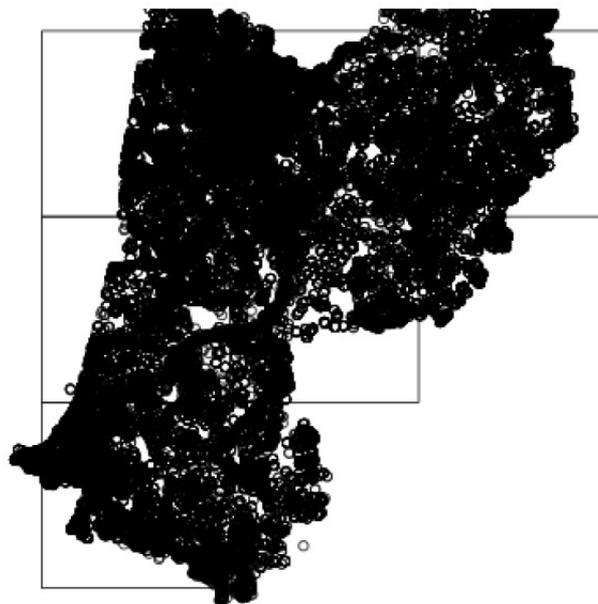
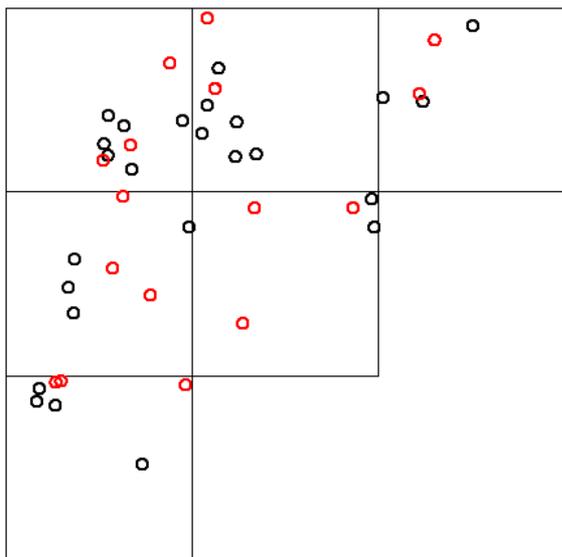
Comparaison des variations d'abondances entre grands carrés

$\frac{\text{STOC2j}}{\text{STOC2j}'}$

$\frac{\text{STOC1j}}{\text{STOC1j}'}$

$\frac{\text{LPOj}}{\text{LPOj}'}$

$\frac{(\text{STOC1+LPO})j}{(\text{STOC1+LPO})j'}$



Comparaison des variations d'abondances entre grands carrés

	STOC1	LPO(*)	(STOC1+LPO)
STOC2	$R^2 = 8.3\%$	$R^2 = 8.8\%$	$R^2 = 26.5\%$

(*) effort = nombre total d'observations par grand carré

- Avec cette méthode, nous pouvons estimer les A_{ij} pour des espèces présentes uniquement dans le jeu de données $K=1$ (effort inconnu).

=> Nouvelle manip, nous enlevons une espèce de STOC1 et on l'estime avec STOC1incomplet+LPO

Comparaison des variations d'abondances entre grands carrés

	(STOC1+LPO)	(STOC1incomplet+LPO)
STOC2	$R^2 = 26.5\%$	$R^2 = 26.3\%$

Rapportage relatif des espèces dans les données opportunistes:

Moineau D	0.19
étourneau	0.22
tourterelle T	0.36
troglodyte	0.40
F à tête noire	0.46
Pinson	0.48
Merle	0.48
corneille	0.49
pigeon ramier	0.53
M charbonnière	0.55
Hirondelle R	0.56

Rougegorge	0.56
M Bleue	0.61
verdier	0.63
P véloce	0.64
Grimpereau des J	0.67
Pic épeiche	0.75
Sittelle	0.80
Geais	0.85
Hypolaïs	0.86
Rougequeue N	1,0

What are the key hypotheses?

$$X_{ijk} \sim \text{Poisson}(A_{ij} \cdot E_{jk} \cdot P_{ik})$$

HYP Effort only depend on site and scheme
 Detectability only depend on species and scheme

=> no special effort toward some species in some sites!!

=> Homogénéité du comportement des observateurs

Perspectives

Introduire une stratification: l'habitat

$$A_{ij} \frac{S_{ih} V_{hj}}{\sum_{h'} S_{ih'} V_{h'j}} \times E_{j2} \frac{q_{h2}}{\sum_{h' \in l} q_{h'2}} \times P_{ihk}$$

Répartition des espèces en fonction de leur préférence S_{ih} et de l'offre en habitat V_{hj}

Biais d'attention q_h des observateurs vers habitat h parmi ceux présent dans lieu-dit l

Données opportunistes, données structurées : quel avenir ?

- La stratégie eBird : documenter systématiquement l'effort (non-contraint) d'observation

- L' (absence) de stratégie Faune-LPO : laisser le maximum de liberté aux observateurs ; croire dans une certaine homogénéité de l'effort ; (combiner avec des données très standardisées)

+ différences fondamentales sur ouverture des données et modèle économique...

Autres pistes :

⇒ Maximiser la qualité des données

(plutôt que le « big data », la « quality data »)

Ex: « j'ai vu 3 papillons blancs dans mon jardin »



Ex: « j'ai vu 3 papillons blancs dans mon jardin »



Protocole : « nbre maximum d'individus vus simultanément »



Ex: « j'ai vu 3 papillons
blancs dans mon jardin »

Protocole : « nbre maximum
d'individus vus simultanément

Assemblage d'espèces :

-2 bleus

-1 paon du jour

-1 petite tortue



Ex: « j'ai vu 3 papillons
blancs dans mon jardin »

Protocole : « nbre maximum
d'individus vus simultanément

Assemblage d'espèces :

-2 bleus

-1 paon du jour

-1 petite tortue

100- 300 m²

Pas d'anti-limace

Petit coin de friche

<100m d'une forêt

Ce qui fait la qualité d'une donnée naturaliste

⇒ **Précision** (compétence de l'observateur)

⇒ **Répétabilité** (respect du protocole)

⇒ **Représentativité** (plan d'échantillonnage,
post- stratification)

Différents observatoires selon les groupes biologiques et les observateurs

VIGIENATURE

Un réseau de citoyens qui fait avancer la science

Grand public



Observatoire
de la biodiversité
des jardins



Oiseaux
des jardins



Observatoire
des bourdons



SPIPOLL
Suivi photographique
des insectes
pollinisateurs



Sauvages
de ma rue



BioLit

Naturalistes



STOC
Suivi temporel
des oiseaux
communs



SHOC
Suivi hivernal
des oiseaux
communs



Vigie-Chiro



STELI
Suivi
temporel
des libellules



Vigie-Flore



STERF
Suivi temporel
des Rhopalocères
de France

Scolaires



Vigie-Nature École

Professionnels



Observatoire
Agricole
de la Biodiversité



PROPAGE
Protocole papillons
et gestionnaires
d'espaces verts

Autres pistes :

⇒ Combiner données de qualité et données avec
preuves : photo numérique, enregistrement
sonores



Suivi Photographique des Insectes POLLinisateurs

Une initiative labellisée



Un observatoire des pollinisateurs ?



- ⊙ **La pollinisation**, un service écologique menacé
- ⊙ La **diversité** est importante
- ⊙ **Des pressions multiples** : réchauffement climatique, artificialisation du territoire, intensité des pratiques agricoles

Un observatoire des pollinisateurs ?



- ◎ De nombreux points d'échantillonnage nécessaires mais très peu de spécialistes
- ◎ Quelques milliers d'espèces pollinisatrices en France !!

L'idée : recourir à la photographie amateur !



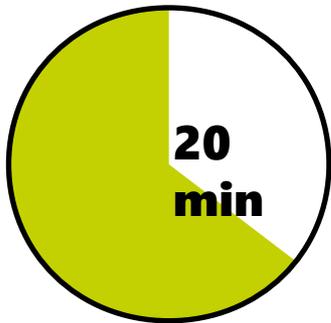


A vous de jouer !

Étape 1

Un safari-photo de 20 minutes





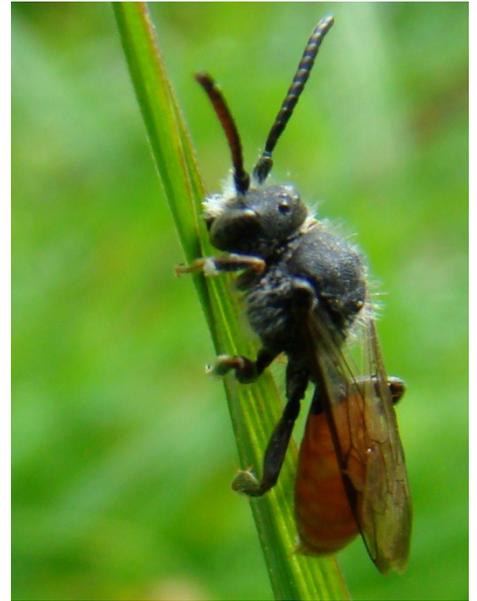


A vous de jouer !

Étape 2

Trier..., recadrer... : collectionner









A vous de jouer !

Étape 3

Trouver le bon nom... parmi 600 possibilités !



Clé d'Identification d'insectes sur photo

19 Descri... Historiqu...

 Longueur des antennes

 Forme des yeux

 Type de coloration du thorax

 Type de coloration de l'abdomen

 Elargissement et pilosité de la patte postérieure

La dernière paire de patte peut être élargie ou non. De même, elle peut être velue ou non.

Elargissement et pilosité de la patte postérieure



Non élargie et peu ou pas velue
(145)

La paire de patte postérieure n'est pas élargie sur toute sa longueur et n'est pas velue.



x

197 Taxons restants (espèces, groupes d'e... Q

L'Abeille Ceratina noire (*Ceratina cucurbitina*)

L'Abeille coucou Epeloides (femelle) (*Epeoloides coecutiens*)

L'Abeille mellifère (*Apis mellifera*)

Ce que propose la clé :



Le syrphe ceinturé



Les chloropides
jaunes



© Philippe Moniotte - Entomopix



A vous de jouer !

Étape 4

**Partager vos collections et commenter celles
des autres participants**





AFFICHER

2010-09-17

Nom de la Fleur: Taxon inconnu de la clé (aster)

Il s'agit d'une fleur : plantée

Il s'agit d'une habitat : jardin privé /

BIVILLE-SUR-MER (76098), SEINE-MARITIME (76),

HAUTE-NORMANDIE (23)

par : étamines

[TOUTES SES COLLECTIONS DANS LES GALERIES](#)



Dernière(s) identification(s):
Les Mouches à damier

AFFICHER



Dernière(s) identification(s):
La Syritte pialante

AFFICHER



Dernière(s) identification(s):
Les Mégachiles rayés

AFFICHER



Dernière(s) identification(s):
Les Mouches aux reflets métalliques

AFFICHER

COMMENTAIRES DES INTERNAUTES

par : cybelle 2011-04-23

Superbe votre collection Etamines, les insectes se bouscuaient pour vous faire plaisir, vous avez dû vous régaler.

par : Fernand 2011-04-23

Très belles photos. Les fleurs sont magnifiques et les insectes également. Beau travail étamines .

Des données accessibles a tous

LES COLLECTIONS

FILTRES

Mes filtres

Entrer un nom pour ce filtre

Enregistrer

PSEUDO

RÉINITIALISER

DATE

RÉINITIALISER

FLEUR

RÉINITIALISER

INSECTE

RÉINITIALISER

CONDITIONS D'OBSERVATION

RÉINITIALISER

LOCALISATION

RÉINITIALISER

Localiser la fleur : utiliser les champs et/ou la carte ci-dessous :

[CHERCHER](#)

(Le nom d'un village, d'une ville, d'une région, d'un département ou un code postal.)

ou

No INSEE. [CHERCHER](#)



[RECHERCHER DES INSECTES](#)

[RECHERCHER DES COLLECTIONS](#)

Des données accessibles a tous

publications [Mathématique...] | Galeries | SPIPOLL | ménage

ALIX-EN-PROVENCE FAISSE ROI! ETTE 25 FEVRIER 2011

ENREGISTRER DANS MES COLLECTIONS PRÉFÉRÉES | PRÉCÉDENT | SUIVANT



ARTICLE

2011-02-25

Nom de la fleur : Taxon inconnu de la clé (Fausse roquette, Diploclavis encucoides)
Il s'agit d'une fleur : spontanée.
Il s'agit d'une habitat : bord de route / ALIX-EN-PROVENCE (13001), BOUCHES-DU-RHÔNE (13), PROVENCE-ALPES-CÔTE D'AZUR (93)
par : Prisca

TOUTES SES COLLECTIONS DANS LES GALERIES



 <p>Commenter l'identification: Les Xylocopes</p> <p>ARTICLE</p>	 <p>Commenter l'identification: L'Abaille matifère</p> <p>ARTICLE</p>	 <p>Commenter l'identification: Les Abailles rayées</p> <p>ARTICLE</p>
 <p>Commenter l'identification: Les Eristalis aux yeux ponctués</p> <p>ARTICLE</p>	 <p>Commenter l'identification: Les Mouches à effluve à dôme noir</p> <p>ARTICLE</p>	 <p>Commenter l'identification: Les Abailles rayées</p> <p>ARTICLE</p>
 <p>Commenter l'identification: Les Syrphes à abdomen fin</p> <p>ARTICLE</p>	 <p>Commenter l'identification: Les Syrphes à l'aspect de bourdon</p> <p>ARTICLE</p>	 <p>Commenter l'identification: Les Eristalis (autres)</p> <p>ARTICLE</p>
 <p>Commenter l'identification: Le Syrphé calénaire</p> <p>ARTICLE</p>		

publications [Mathématique...] | Galeries | SPIPOLL | ménagerie de verre

LES XYLOCOPES

SUIVANT | RETOUR À LA COLLECTION →



IDENTIFICATION

Les Xylocopes par : Prisca 2011-02-28

Ancienne identification

ÉMETTRE UN DOUTE SUR L'IDENTIFICATION

INFORMATIONS COMPLÉMENTAIRES

Date: 2011-02-25
Heure: 11:50 à 12:50
Ciel - couverture nuageuse: 0-25%
Température: 10-20°C
Vent: faible, irrégulier
Fleur à l'ombre: Non

COMMENTAIRES DES INTERNAUTES

Ajouter un commentaire

Aucun commentaire enregistré.

Des données validées

LES HALICTES (FEMELLES)

[RETOUR À LA LISTE](#) → [RETOUR À LA COLLECTION](#) →



Alexis Borges
Lepidoptera



Mathieu de Flores
Arthropoda



Serge Gadoum
Hymenoptera, Syrphidae



Hervé Guyot
Arthropoda



Pierre Zagatti
Coleoptera



IDENTIFICATION

Les Halictes (femelles) par : mathieu.opie 2010-11-08

Cette identification a été effectuée par un expert. Elle est considérée comme valide.

Ancienne identification

2010-08-10 : Les Abeilles difficiles à déterminer <Apidae et autres>, L'Abeille mellifère <Apis mellifera>, L'Abeille à culottes (femelle) <Dasypoda hirtipes>, Les Abeilles rayées <Andrenidae et autres> par : Nicolas

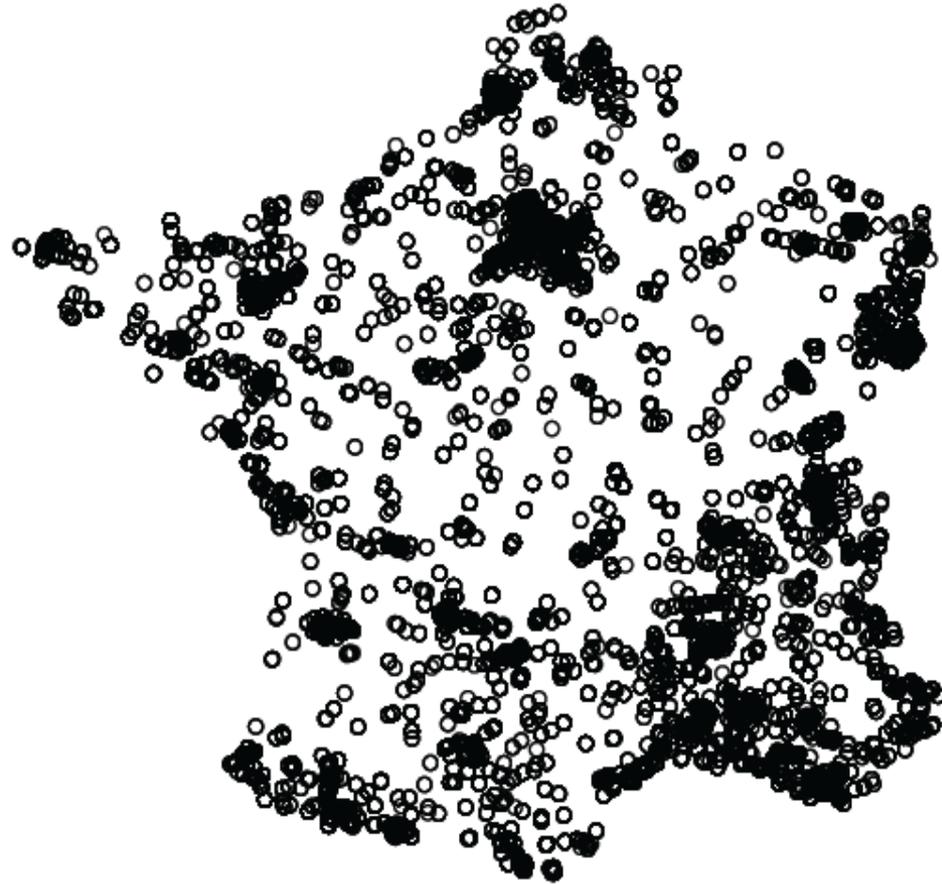
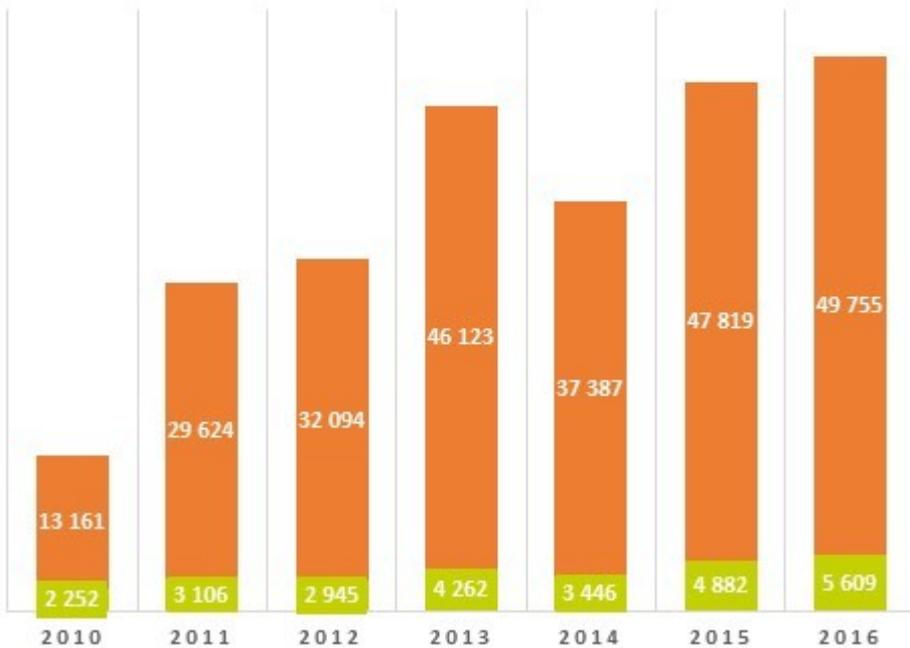


Quelques chiffres sur la participation

Des données en quantité

EVOLUTION DU NOMBRE DE
DONNÉES

■ Collections ■ Photos



Fin 2016,



LE SPIPOLL, C'EST



1 364
OBSERVATEURS



310 233 PHOTOS
AU TOTAL

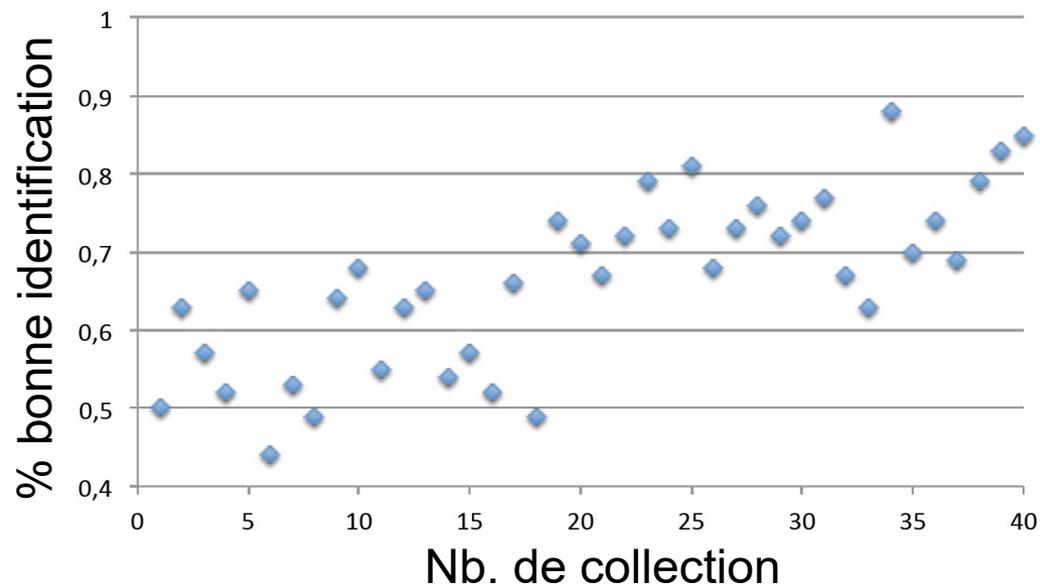
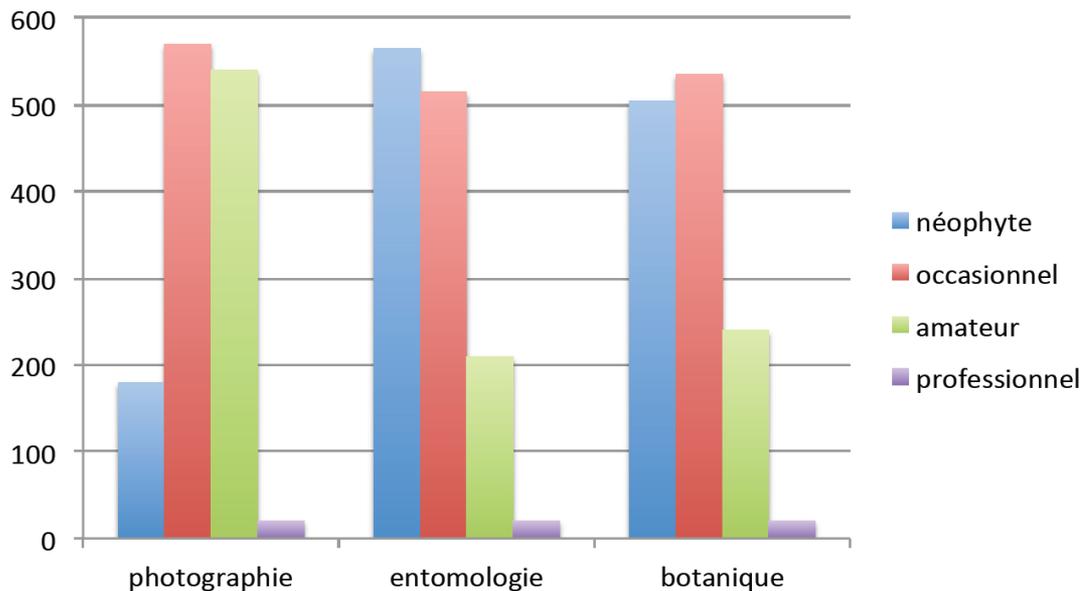


14 307 HEURES
D'OBSERVATION
soit 1 670 journées
de 7 heures



2 701 COMMUNES
AVEC AU MOINS
UNE COLLECTION

Une porte d'entrée sur l'entomologie



SPIPOLL



Des partenaires complémentaires



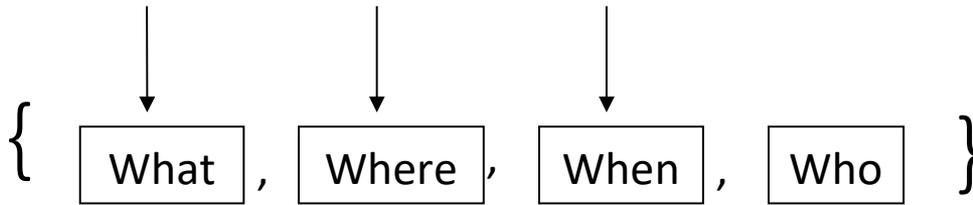
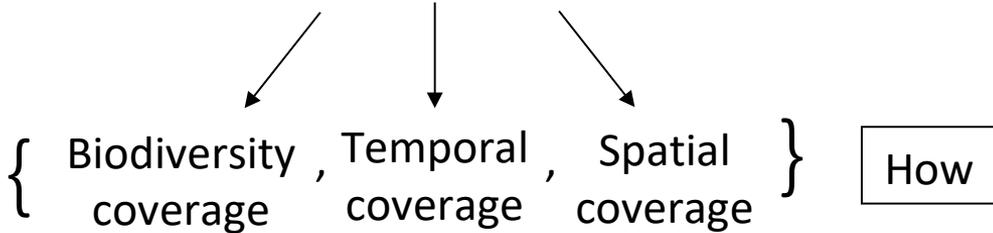
www.spipoll.org



Merci !

Why

Set of rules and constraints



Incidental data

Project-based citizen science

(semi) Structured data with minimum information for monitoring biodiversity changes